ELSEVIER

Analytica Chimica Acta 346 (1997) 307–318

ANALYTICA CHIMICA ACTA

Assessment of new constraints applied to the alternating least squares method

A. de Juan[a,b,*], Y. Vander Heyden[b], R. Tauler[a], D.L. Massart[b]

[a]*Departament de Química Analítica, Universitat de Barcelona, Diagonal 647, 08028 Barcelona, Spain*
[b]*ChemoAC, Farmaceutisch Instituut, Vrije Universiteit Brussel, Laarbeeklaan 103, B-1090 Brussels, Belgium*

Received 6 December 1996; received in revised form 6 December 1996; accepted 16 December 1996

Abstract

The introduction of constraints in multivariate curve resolution methods, such as the Alternating Least Squares (ALS), is commonly used to limit the span of possible solutions, guiding the iterative process to a final result as close as possible to the true situation. In the present work, two modifications of the unimodality constraint and a new constraint for chromatographic concentration profiles related to the prevention of fronting have been checked. Simulated data sets as well as real data have been used to evaluate the effect of these new constraints in the resolution results. The parameters measured to assess the goodness of the constraints are related to the recovery of the concentration profiles and to the quality of the data fit.

Keywords: Chromatography; Constraints; Curve resolution; Alternating least squares; Unimodality

1. Introduction

The application of constraints is a common operation employed in iterative curve resolution methods, such as the Alternating Least Squares (ALS) [1–3], to narrow the span of feasible solutions to those chemically meaningful [4–6].

In a wide sense, the concept of constraint would include any general feature of the data sets translated into mathematical language. However, a careful attitude has to be adopted in order to avoid false generalizations (e.g. not all the chromatographic peaks have a Gaussian shape) and the researcher has to be aware that the effectiveness of a constraint can be strongly affected by the way it has been implemented.

Within the family of constraints, selectivity is the most important to resolve any kind of data sets [1]. If this feature exists for all the compounds, the ambiguity associated with the factor analysis decomposition of bilinear matrices disappears and unique solutions can be attained [1,8]. However, having partial or null selectivity is the most frequent situation in real data sets. When the latter occurs, the role of some other kind of constraints related to the chemical features of concentration profiles and to the instrumental responses acquires relevance and helps to decrease significantly the domain of possible solutions. The development of new constraints belonging to this last group becomes then specially interesting for the resolution of the most complex data sets.

Numerous works have reported the usefulness of constraints for curve resolution methods like non-

*Corresponding author. E-mail:annaj@zeus.qui.ub.es.

PII S0003-2670(97)00145-1

negativity (applied to both concentration profiles and instrumental responses) [1–7,9–14], unimodality (i.e. presence of an only maximum in each concentration profile) [1–3.5–7,10–14] and closure (i.e. the total amount of the reactive constituents is forced to be the same in all the stages of a process) [1,2,10–13,15,16]. In the present work, two modifications of the unimodality constraint are examined, as well as a new constraint for chromatographic concentration profiles (called hereafter symmetry) related to the prevention of front-tailed peaks.

A two-level full factorial design has been used to generate simulated data sets which could yield a reliable picture of the effect of these new constraints in the resolution procedure. The responses measured to assess the goodness of the constraints are related to the recovery of the qualitative information and to the quality of the data fit. All conclusions inferred from this basic study have been afterwards confirmed working with a real example.

2. Theory

2.1. Brief description of the ALS method

The ALS belongs to the family of iterative curve resolution methods. All these techniques share a general working procedure consisting of the refining of some initial estimates (either chromatographic or spectral) by using constraints related to the intrinsic features of the data (i.e. selectivity, zero-concentration regions,...) or to the chemical characteristics of the experimental system (i.e. non-negative concentration profiles or spectra, unimodality,...). In the present work, the initial estimates to be input in the ALS method are always built in the chromatographic direction applying the results obtained in the needle algorithm [17]; then, a constrained alternating least-squares optimization procedure runs till the convergence criterion is reached. The matrices of the concentration profiles and spectra obtained in the last iterative cycle are the definitive solutions of the resolution method when convergence is achieved. A more detailed explanation of the ALS method is out of the scope of this paper and can be found in previous works of the authors [1–3,10–14].

2.2. Presentation of the new constraints

2.2.1. Horizontal unimodality

Equal in concept to the classical unimodal constraint (i.e. no more than one maximum is allowed), this variety differs from the original version in the way it is implemented.

The common steps in both implementations can be summarized as follows:

1. Location of the largest maximum in the concentration profile (m).
2. Suppression of the left local maxima.
3. Suppression of the right local maxima.

The difference lies in the elimination of the secondary maxima: the classical unimodality (a) sets the non-unimodal elements equal to zero and the new implementation (b) equals these elements to the nearest element keeping the unimodal condition. In algorithmic notation, steps 2 and 3 can be expressed as:

2. if $c(m-i) > r \times c(m-i+1)$

(a) $c(m-i) \approx 0$ (for classical unimodality)

(b) $c(m-i) = c(m-i+1)$

(for horizontal unimodality)

3. if $c(m+i+1) > r \times c(m+i)$

(a) $c(m+i+1) \approx 0$ (for classical unimodality)

(b) $c(m+i+1) = c(m+i)$

(for horizontal unimodality)

where $c(m)$ is the maximum value of the concentration profile and the pairs $[c(m-i),\ c(m-i+1)]$ and $[c(m+i+1),\ c(m+i)]$ are the consecutive concentration values to be compared when looking for left and right local maxima, respectively. The parameter r can be optionally larger than 1; if this is the case, small departures of the unimodality are accepted.

From a graphical point of view, the classical constraint cuts the non-unimodal part of the concentration profile *vertically*, whereas the new modification does it *horizontally*. Fig. 1 illustrates the effect of both varieties of unimodality on a concentration profile. The plot stresses the positive behaviour of the new implementation for noisy peaks, normally related to minor compounds. Such peaks contain noisy spikes that are

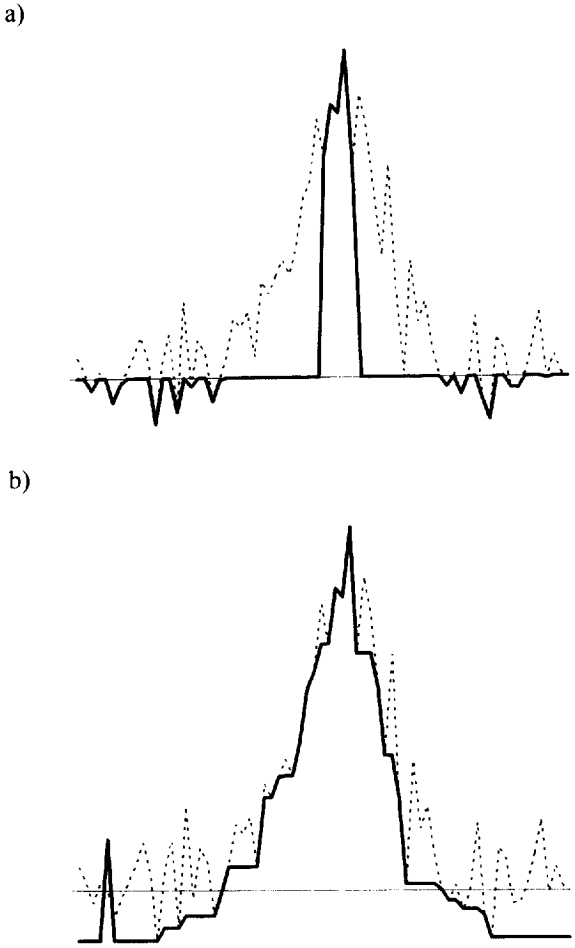

Fig. 1. Effect of the two implementations of the unimodal constraint on a concentration profile: (a) classical unimodality; (b) horizontal unimodality. Dotted line: original peak. Solid line: constrained peak.

detected as secondary maxima, this fact leading to the wrong suppression of a big part of the concentration profile when the classical implementation is applied. In contrast, the horizontal elimination of these maxima allows to keep a constrained concentration profile with a shape much closer to the original peak.

2.2.2. *Localized unimodality*

This constraint has to be considered as a more demanding version of the normal unimodality, whatever implementation applied. Both the existence of an only maximum and its position in the concentration profile are controlled. The positions taken as reference maxima for the different compounds are previously determined by using methods like the needle algorithm [17], OPA [18] or SIMPLISMA [9].

The application sequence of this constraint is detailed below. The text in italics indicates the additional step, absent in the normal unimodality, for which the algorithmic notation is also included.

1. Location of the largest maximum in the concentration profile (m).
2. *Comparison between the position of the reference maximum (mt) and the position obtained from the resolution profiles (m). Relocation of the peak maximum if necessary*

   if $abs(m - mt) > rpeak$

   $m = mt$.
3. Suppression of the left local maxima.
4. Suppression of the right local maxima.

The maximum shift allowed in the peak position is represented by the parameter *rpeak*, whose value is selected by the user. Analogously to the other varieties of this constraint, small departures of the unimodal condition can be optionally accepted. Steps 3 and 4 have been applied as in the horizontal unimodality.

2.2.3. *Symmetry*

Despite its name, the present constraint does not transform all the concentration profiles into symmetrical signals. Tailed peaks are accepted, provided that this asymmetry in the profile is placed after the peak maxima. Thus, real chromatographic situations where column ageing or other causes produce the distortion of the theoretical Gaussian peaks can be correctly reproduced.

The symmetry constraint is focused on the suppression of front-tailed peaks. When such a peak shape arises from the resolution results, the concentration profile of the affected compound is forced to be symmetrical. Although phenomena of band fronting can occasionally be found in real data, their occurrence is far less common than the appearance of band tailing and it is practically reduced to the domain of ion-pair chromatography [19]. Visual evidence of band fronting in a chromatogram would indicate the

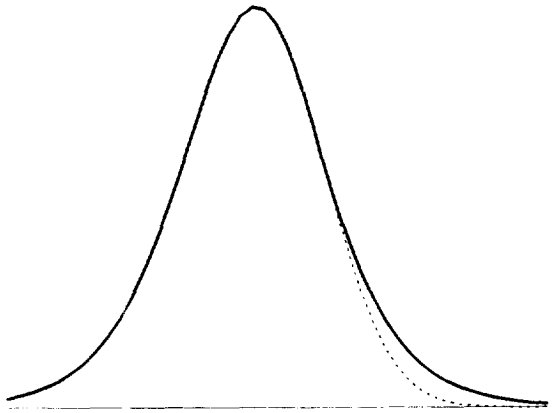

Fig. 2. Effect of the symmetry constraint on a concentration profile. Dotted line: original peak; solid line: constrained peak.

need of a modification in the separation parameters (i.e. increase of temperature, decrease of the amount of injected sample,...). The non-adequacy of the symmetry constraint would be limited to those rare situations where band fronting appears and cannot be chromatographically suppressed. Therefore, the proposal of such a constraint is chemically reasonable in by far the most cases.

The symmetry constraint reshapes the front-tailed peaks by making the back half of the peak symmetrical to the front (see Fig. 2). Gaussian and tailed peaks are not modified. The mathematical formulation of this constraint is very simple and can be explained in two steps:

1. Location of the largest maximum in the concentration profile (m).
2. Detection and suppression of band fronting:

$$\text{if } c(m-i) > r \times c(m+i)$$
$$c(m+i) = c(m-i)$$

where the pair $[c(m-i),\ c(m+i)]$ represents two concentration values equidistant to the peak maxima. Band fronting is present when the values in the left half of the peak are bigger than those in the right half of the peak. The parameter r can be optionally bigger than 1; departures of the symmetry constraint are then accepted.

3. Data sets

3.1. Simulated data sets

Two-compounds simulated data sets have been used to assess the constraints presented above. The choice of this kind of systems is related to the role they play as reference models in peak purity problems and to the fact that many real multicompound samples can be resolved analyzing submatrices of peak clusters which do not contain usually more than two or three overlapped substances. Furthermore, the simple structure of these systems allows a clear interpretation of the effects of the constraints tested, in contrast to the vague conclusions inferred when very complex simulations are employed.

In all the examples the two peaks are slightly tailed and there is a major and a minor compound. Severely overlapped spectra have been chosen on purpose for all the simulated data sets (see Fig. 3). Since the constraints to be assessed affect the concentration profiles, a better evaluation of their real effect can be carried out when no spectral selectivity can influence the resolution procedure.

The general validity of the conclusions related to the quality of the new constraints has been ensured through the non-arbitrary selection of the simulations employed in the testing procedure. A two-level factorial design has been used to determine the features of the generated data sets [20]. Table 1 shows the relevant information concerning the factors and their

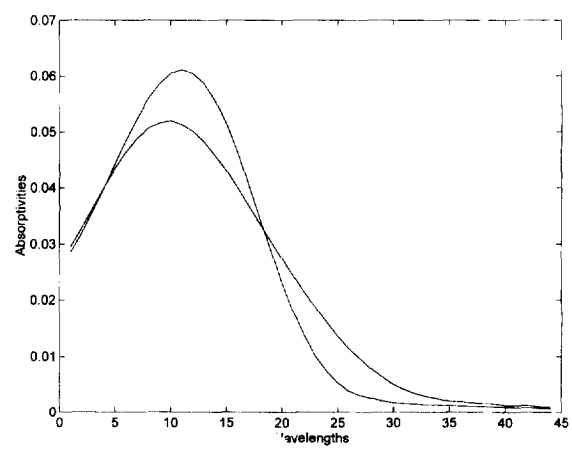

Fig. 3. Spectra used in the simulated data sets.

Table 1
Two-level full factorial design used to generate the simulated data sets. List of the coded properties related to each simulation

Factors	levels	
	(−)	(+)
Constraint (A)	Absence	Presence
Resolution (B)	0.2	0.8
Ratio minor/major compound (C)	1 : 100	1 : 10
Noise pattern (D)	Homoscedastic	Heteroscedastic
S/N ratio minor compound	20	50

Simulation	Factors				
	A	B	C	D	E
1	−	−	−	−	−
2	+	−	−	−	−
3	−	+	−	−	−
4	+	+	−	−	−
5	−	−	+	−	−
6	+	−	+	−	−
7	−	+	+	−	−
8	+	+	+	−	−
9	−	−	−	+	−
10	+	−	−	+	−
11	−	+	−	+	−
12	+	+	−	+	−
13	−	−	+	+	−
14	+	−	+	+	−
15	−	+	+	+	−
16	+	+	+	+	−
17	−	−	−	−	+
18	+	−	−	−	+
19	−	+	−	−	+
20	+	+	−	−	+
21	−	−	+	−	+
22	+	−	+	−	+
23	−	+	+	−	+
24	+	+	+	−	+
25	−	−	−	+	+
26	+	−	−	+	+
27	−	+	−	+	+
28	+	+	−	+	+
29	−	−	+	+	+
30	+	−	+	+	+
31	−	+	+	+	+
32	+	+	+	+	+

levels, the selected design and the coded properties of all the simulated data sets.

The factors included in the design are parameters whose influence on the results of a resolution method is either proven or at least potential and the two levels set for each of the factors have been chosen trying to cover a wide span of real situations. Thus, the constraint to be checked is introduced as a factor, whose qualitative levels are absence (−) and presence (+). The negative level of this factor has been differently defined according to the constraint to be checked (detailed explanations in Section 4.1.2 and next). The remaining factors are features of the chromatographic system, namely the resolution between peaks, the concentration ratio between major and minor compound, the noise pattern and the signal-to-noise ratio for the minor compound. The heteroscedastic noise has been simulated by adding to a homoscedastic background a scaled contribution of this base noise proportional to the square root of the intensity of the signal, i.e. for each *ij*th point of the data matrix, the noise added can be described as follows:

$$\text{total noise}_{ij} = (\text{homosc. noise})_{ij} + \sqrt{\text{noise-free signal}_{ij}} \times (\text{homosc. noise})_{ij}. \quad (1)$$

The concept of noise level has been represented through the signal-to-noise ratio for the minor compound instead of using the common measurement % of noise with respect to the largest absorbance. In systems containing major and minor compounds, the *S/N* ratio for the small species has been considered more informative than a general parameter to evaluate the distortion of the minor signals, strongly tied to the possibility of resolving these compounds. For both homoscedastic and heteroscedastic systems, the signal-to-noise ratio for the minor compound follows the expression below:

$$S/N = \frac{\max(\text{noise-free signal minor compound})}{\sqrt{\frac{\sum_{ij}(\text{homosc. noise})_{ij}}{i \times j}}}, \quad (2)$$

where the noise-free signal for the minor compound comes from the outer product $c_m \times s_m$, where c_m and s_m are the simulated concentration profile and spectrum associated with this constituent, respectively. Please note that the apparently large *S/N* values used in the present work are actually the maxima allowed for this parameter in the data sets, since far from the position of the signal maximum the ratio between

signal and noise becomes considerably lower. Variations of the minor/major concentration ratios have been simulated keeping fixed the signal related to the minor compound and modifying the signal of the major compound appropriately. Thus, all the simulations having the same *S/N* ratio for the minor compound have the same amount of added noise as well, since the noise-free signal for the minor compound remains invariant. This strategy simplifies the simulation process and the further interpretation of the results.

3.2. Real data sets

Two-compound data sets, reported as reference systems for peak purity studies [21], are used to confirm the conclusions inferred from the simulated data about the goodness of the checked constraints. These real systems contain hydrocortisone as major compound and prednisone as minor constituent. The spectra of these species are the same used in the simulated data sets, shown in Fig. 3.

4. Results and discussion

4.1. General remarks

To test each of the constraints, the ALS method has been applied to the series of 32 experiments designed according to the information in Table 1. ALS has always been run forcing non-negative concentration profiles and spectra. Even though the use of selectivity is essential in this resolution method, the information regarding this point has not been taken into account in any of the systems analyzed. In the spectral direction, no selectivity can be found as pointed out above, whereas in the chromatographic direction the detection of this feature is only evident in some systems where the resolution between peaks is large (Rs=0.8) and the signal-to-noise ratio for the minor compound quite favourable. The application of selectivity in these last cases would be unquestionable in a normal analysis; no explicit use of this constraint has been done in the present study because in the cases where it is present its strong effect would mask the influence in the ALS solution coming from the constraints to be checked.

The responses collected to analyze the effects of the proposed constraints in the quality of the resolution results are related to the recovery of the qualitative information (i.e. dissimilarities between actual and recovered concentration profiles) and to the error associated with the definitive solution (i.e. standard deviation of the residuals, σ, and the lack of fit). The mathematical expressions associated with the responses are shown below:

$$\text{dissimilarity} = \sqrt{1 - (\text{correlation coefficient})^2},$$

$$\sigma = \sqrt{\frac{\sum_{ij}(d_{ij} - d_{ij}^*)^2}{i \times j}},$$

$$\text{lack of fit} = \sqrt{\frac{\sum_{ij}(d_{ij} - d_{ij}^*)^2}{\sum_{ij} d_{ij}^2}},$$

where d_{ij} are the experimental data and d_{ij}^* the reproduced data by using the ALS optimization. Subscripts i and j are referred to rows and columns of the data matrix, respectively.

Tables 2–4 list the numerical responses obtained in the study of each of the constraints tested. These data have been used to calculate the main effects of the factors and the 2, 3 and 4-factor interactions [20]. The information associated with the latter calculations is graphically presented in the normal probability plots shown in Figs. 4–6 [20]. These graphs represent the value of each of the calculated effects vs. the expected probability it should have if all the effects were normally distributed. The vertical scale in the plots has been transformed, so that the plotted effects could be fitted with a straight line when changes in the levels of all the factors considered do not cause any noticeable variation in the responses or, if they do, this variation is random and does not follow any definite tendency. Positive or negative points falling far from this line indicate either factors or interactions having an important influence on the responses. The graphical conclusions concerning the significance of the effects have been statistically confirmed through the application of a *t*-test in which it is investigated whether or not an effect is significantly different from zero [20]. To do so, the numerical value of each potentially significant effect is compared with a critical effect value, calculated as the product between the averaged value

Table 2
Responses obtained from the ALS runs used to test the horizontal unimodality

Simulations	Responses			
	Dis(c1)[a]	Dis(c2)[b]	σ	Lack of fit (%)
1	1.60E-01	5.26E-01	0.005	6.31
2	1.49E-01	2.09E-01	0.0023	3.06
3	2.51E-03	4.35E-01	0.0004	0.6
4	2.18E-03	1.38E-01	0.00027	0.37
5	2.31E-01	4.07E-01	0.0007	8.6
6	1.10E-01	1.72E-01	0.00034	4.2
7	3.23E-02	4.15E-01	0.00037	4.9
8	3.52E-02	1.55E-01	0.00026	3.45
9	2.21E-01	7.78E-01	0.0098	13
10	2.21E-01	3.09E-01	0.0055	7.31
11	3.76E-03	3.25E-01	0.00047	0.62
12	3.31E-03	1.52E-01	0.00035	0.48
13	6.91E-01	6.55E-01	0.0042	52.1
14	3.44E-01	4.61E-01	0. 0016	19.86
15	6.77E-02	7.78E-01	0.00082	10.8
16	7.16E-02	3.14E-01	0.00057	7.57
17	1.12E-01	2.47E-01	0.0012	1.63
18	1.06E-01	7.96E-02	0.0006	0.79
19	1.17E-03	8.02E-02	0.00012	0.16
20	1.19E-03	6.72E-02	0.0001	0.16
21	1.55E-01	7.05E-02	0.00028	3.5
22	1.29E-01	1.65E-01	0.00014	1.71
23	1.04E-02	8.91E-02	0.00013	1.7
24	1.01E-02	6.68E-02	0.00011	1.54
25	1.31E-01	1.97E-01	0.0018	2.4
26	1.49E-01	1.30E-01	0.0012	1.56
27	2.36E-03	8.04E-02	0.00019	0.26
28	1.82E-03	7.82E-02	0.00016	0.21
29	7.35E-01	6.47E-01	0.0027	33
30	2.78E-01	1.82E-01	0.00044	5.41
31	2.28E-02	1.34E-01	0.00025	3.37
32	2.34E-02	9.11E-02	0.00022	2.9

[a] Dissimilarity between the actual and the recovered profile for the minor compound.
[b] Dissimilarity between the actual and the recovered profile for the major compound.

Table 3
Responses obtained from the ALS runs used to test the localized unimodality

Simulations	Responses			
	Dis(c1)[a]	Dis(c2)[b]	σ	Lack of fit (%)
1	2.09E-01	1.49E-01	0.0023	3.06
2	2.09E-01	1.49E-01	0.0023	3.06
3	1.38E-01	2.18E-03	0.00027	0.37
4	1.38E-01	2.18E-03	0.00028	0.37
5	1.72E-01	1.10E-01	0.00034	4.2
6	1.72E-01	1.10E-01	0.00034	4.22
7	1.55E-01	3.52E-02	0.00026	3.45
8	1.55E-01	3.52E-02	0.00026	3.45
9	3.09E-01	2.21E-01	0.0055	7.31
10	4.28E-01	2.13E-01	0.0056	7.51
11	1.52E-01	3.31E-03	0.00035	0.48
12	1.52E-01	3.31E-03	0.00035	0.48
13	4.61E-01	3.44E-01	0.0016	19.86
14	3.22E-01	3.44E-01	0.0017	21.18
15	3.14E-01	7.16E-02	0.00057	7.57
16	3.14E-01	7.16E-02	0.00057	7.57
17	7.96E-02	1.06E-01	0.0006	0.79
18	7.96E-02	1.06E-01	0.00059	0.79
19	6.72E-02	1.19E-03	0.0001	0.16
20	6.72E-02	1.19E-03	0.00012	0.16
21	1.65E-01	1.29E-01	0.00014	1.71
22	1.65E-01	1.29E-01	0.00014	1.71
23	6.68E-02	1.01E-02	0.00011	1.54
24	6.68E-02	1.01E-02	0.00011	1.54
25	1.30E-01	1.49E-01	0.0012	1.56
26	1.30E-01	1.49E-01	0.0012	1.56
27	7.82E-02	1.82E-03	0.00016	0.21
28	7.82E-02	1.82E-03	0.00016	0.21
29	1.82E-01	2.78E-01	0.00044	5.41
30	1.85E-01	2.61E-01	0.00044	5.5
31	9.11E-02	2.34E-02	0.00022	2.9
32	1.34E-01	2.28E-02	0.00022	2.9

[a] Dissimilarity between the actual and the recovered profile for the minor compound.
[b] Dissimilarity between the actual and the recovered profile for the major compound.

corresponding to the 3-factor interaction effects (such higher-order interactions are assumed to measure differences arising from experimental error [20]) and the t-value corresponding to a 95% significance level and a number of degrees of freedom equal to the number of 3-factor interactions in the design. When the value of an effect is larger than the critical reference value, the tested effect is found to be significant.

In all normal probability plots, the main effect caused by the constraint is identified with the letter A. Effects found to be significant through both visual inspection and statistical diagnostic are also labelled with their corresponding capital letters (see Table 1).

The improvement of the ALS results is always connected with a decrease in the numerical value of the responses. Indeed, better recoveries of the qualitative information are reached when the dissimilarities

Table 4
Responses obtained from the ALS runs used to test the symmetry constraint

Simulations	Responses			
	Dis(c1)[a]	Dis(c2)[b]	σ	Lack of fit (%)
1	2.09E-01	1.49E-01	0.0023	3.06
2	1.41E-01	3.39E-01	0.0069	9.2
3	1.38E-01	2.18E-03	0.00027	0.37
4	1.38E-01	2.18E-03	0.00027	0.37
5	1.72E-01	1.10E-01	0.00034	4.2
6	1.72E-01	1.10E-01	0.00034	4.2
7	1.55E-01	3.52E-02	0.00026	3.45
8	1.46E-01	4.00E-02	0.0003	4.02
9	3.09E-01	2.21E-01	0.0055	7.31
10	2.94E-01	2.53E-01	0.0066	8.72
11	1.52E-01	3.31E-03	0.00035	0.48
12	2.59E-01	3.34E-03	0.0004	0.53
13	4.61E-01	3.44E-01	0.0016	19.86
14	4.72E-01	3.24E-01	0.0029	13.9
15	3.14E-01	7.16E-02	0.00057	7.57
16	2.50 E-01	8.15E-02	0.00086	8.69
17	7.96E-02	1.06E-01	0.0006	0. 79
18	7.41E-02	2.02E-01	0.0023	3.04
19	6.72E-02	1.19E-03	0.0001	0.16
20	1.07E-01	1.27E-03	0.00014	0.19
21	1.65E-01	1.29E-01	0.00014	1.71
22	1.65E-01	1.29E-01	0.00014	1.7
23	6.68E-02	1.01E-02	0.00011	1.54
24	6.88E-02	1.13E-02	0.00011	1.53
25	1.30E-01	1.49E-01	0.0012	1.56
26	1.14E-01	3.58E-01	0.0065	8.59
27	7.82E-02	1.82E-03	0.00016	0.21
28	7.22E-02	1.82E-03	0.00016	0.21
29	6.47E-01	7.35E-01	0.0027	33
30	2.38E-01	3.47E-01	0.00055	6.75
31	1.34E-01	2.28E-02	0.00025	3.37
32	1.34E-01	3.94E-02	0.00032	4.22

[a] Dissimilarity between the actual and the recovered profile for the minor compound.
[b] Dissimilarity between the actual and the recovered profile for the major compound.

between the actual and the calculated concentration profiles become lower and achievements of more accurate fits occur when the error parameters diminish.

Factors or interactions showing significant negative effects indicate an improvement of the ALS results when going from the negative level to the positive level in the simulations. On the contrary, significant positive effects are associated with factors or interactions causing improvements in the ALS results when going from the positive to the negative level. Therefore, the interpretation of the results have to be performed according to the chemical meaning of

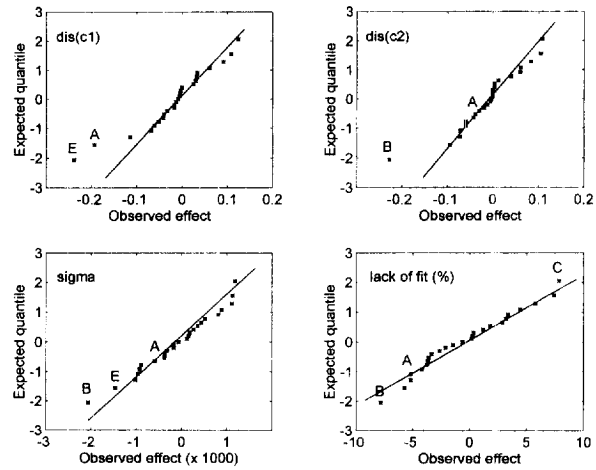

Fig. 4. Normal probability plots related to the assessment of the horizontal unimodality constraint. Significant effects are labelled with capital letters (see Table 1 for identification) and the constraint effect is always identified with the letter A. The responses analyzed are the dissimilarity between the actual and the recovered concentration profile for the minor compound, dis(c1), and the major compound, dis(c2), σ and the lack of fit.

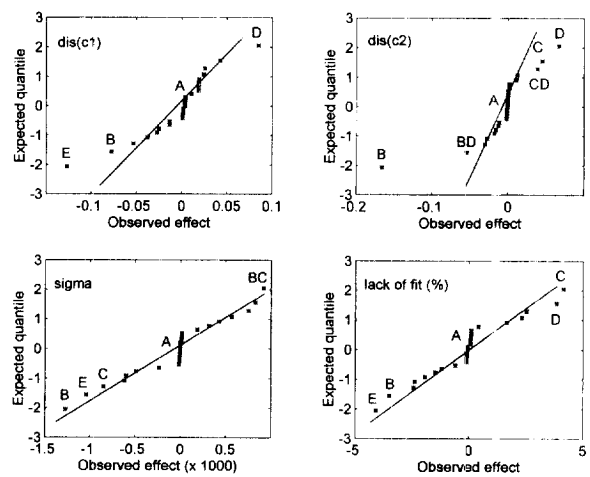

Fig. 5. Normal probability plots related to the assessment of the localized unimodality constraint. Significant effects are labelled with capital letters (see Table 1 for identification) and the constraint effect is always identified with the letter A. The responses analyzed are the dissimilarity between the actual and the recovered concentration profile for the minor compound, dis(c1), and the major compound, dis(c2), σ and the lack of fit.

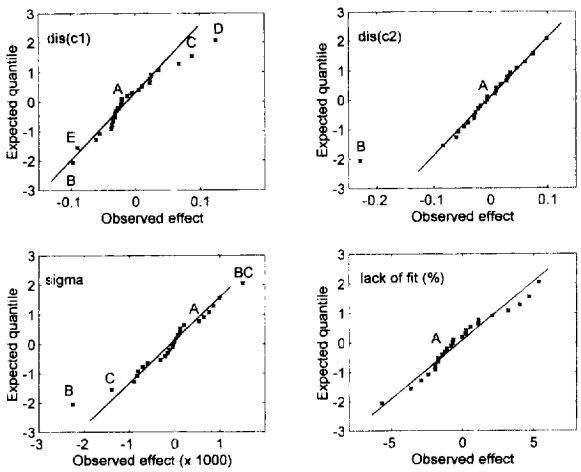

Fig. 6. Normal probability plots related to the assessment of the symmetry constraint. Significant effects are labelled with capital letters (see Table 1 for identification) and the constraint effect is always identified with the letter A. The responses analyzed are the dissimilarity between the actual and the recovered concentration profile for the minor compound, dis(c1), and the major compound, dis(c2), σ and the lack of fit.

the factors and the definition of their levels in the experimental design.

4.1.1. Comments about the data features effects on the ALS results

Figs. 4–6 show some general trends concerning the effects of the factors related to the data features on the quality of the ALS results. As pointed out above, these factors are e resolution between peaks, the minor/major concentration ratio, the noise pattern and the *S/N* ratio for the minor compound and they are identified with the capital letters B, C, D and E (see Table 1), respectively.

As it is shown in the lately mentioned figures, the recovery of the qualitative information is positively influenced by increases in the resolution between peaks and in the signal-to-noise ratio for the min- compound (the first factor being clearly the most important in the shape modelling of the major compound and the second in the modelling of the minor). Bigger dissimilarities between actual and recovered concentration profiles appear when the noise pattern is heteroscedastic and when the minor/major concentration ratio in the binary system increases. The negative action of the latter factor is linked with the decrease of the major compound signal in the simulated data sets. The consequent diminution of the signal-to-noise ratio for this constituent and the comparable contribution of both minor and major signals lead globally to a more merged and noisy binary system, where the distinction and correct modelling of the compounds becomes more difficult.

Most of the normal probability plots related to dissimilarities can be interpreted taking into account only the main effects of the factors. However, a comment related to the interactions resolution-noise pattern (BD) and minor/major ratio-noise pattern (CD) is included, since the effects of these combinations have been found significant in the dissimilarity of the major compound in Fig. 5. Both interactions show that variations in the chromatographic resolution (B) and in the minor/major concentration ratio (C) affect differently the quality of the ALS results when occurring in systems with homoscedastic or with heteroscedastic noise pattern. Thus, the BD interaction implies that a decrease in the resolution between peaks is more critical in the recovery of the concentration profile of the main compound when the noise is heteroscedastic, whereas the CD interaction tells that decreases in the proportion of the major compound in the system affect more strongly the modelling of the concentration profile of this compound in the presence of heteroscedasticity. In both cases the explanation is directly related to the stronger diminution of the *S/N* ratio for the major compound induced by the presence of heteroscedasticity when the resolution decreases (BD) and when the proportion of major compound decreases (CD). An examination of Eq. (1) allows to understand the intense effect of the heteroscedastic pattern in both kind of interactions. This expression consists of two terms, the first one including a homoscedastic background (added to all systems, whatever noise pattern they have) and the second involving properly the heteroscedastic contribution. In data sets with equal peaks and different resolutions, the term referred to the homoscedastic contribution remains invariant, whereas the second term, scaled according to the square root of the signal, increases locally in systems with low resolution because the global signal becomes larger due to the big overlap between compounds. The comparison between systems with different minor/major concentration ratios shows that the

variation of the total noise added to the data sets when going from the low level to the high level of this factor is also larger in systems with heteroscedastic noise pattern because of the contribution of the second term in Eq. (1).

A fast examination of the normal probability plots related to the error parameters (and lack of fit) shows that the significant effects are less pronounced in these responses. The smaller variability of the error parameters in the different ALS runs analyzed is simply explained because of the rotational ambiguity associated with the decomposition of the bilinear matrices when selective information is not available (i.e. many products between matrices of concentration profiles, whose columns are linear combinations of the actual profiles, and spectra matrices, whose rows are linear combinations of the actual spectra, can reproduce the original data matrix with a similar fit) [1]. The latter statement reveals the dissimilarities as a more sensitive indicator of changes in the ALS results and therefore, simulated studies for which actual and recovered profiles are available constitute an advisable starting point to assess the influence of any modification in a resolution technique onto the final solutions. In spite of this, the error parameters are the only ones that can be determined when studying real systems and this is the reason why the knowledge of the effects caused by the different factors on them is necessary. In agreement with the dissimilarity studies, the resolution between peaks is the factor having the clearest negative effect (i.e. an increase in the chromatographic resolution reduces the error parameters in the final results); the *S/N* ratio for the minor compound affects in the same sense as well, whereas the existence of a heteroscedastic pattern yields to a worsening of the results. Variations in the minor/major concentration ratio influence differently the σ and the lack of fit. The more merged and noisy binary system arisen from the decrease in the major signal causes an expected worsening in the lack of fit. The apparently inconsistent improvement of the residual standard deviation (σ) is a consequence of the absolute character of this parameter. The values of the residuals, as such, are taken to be averaged. This causes that small residuals associated with small values of the original data matrix appear to be better than slightly bigger residuals associated with much larger numerical values of the original data matrix. This fact proofs the danger of analyzing the residuals without taking into account the magnitude of the elements of the original data matrix they are associated with. The rest of statistically significant effects are less important than those mentioned above and do not deserve an exhaustive explanation.

4.1.2. Comments on the assessment of the tested constraints

4.1.2.1. Horizontal unimodality. The ALS method has been run forcing classical unimodality for the concentration profiles in the experiments with negative constraint level. Horizontal unimodality has been applied instead when the constraint level is positive. Small departures of the unimodal constraint have been allowed in both implementations (tolerance parameter, r=1.1).

Fig. 4 shows the normal probability plots related to the assessment of the horizontal unimodality constraint. The letter A marks the point associated with the effect of this constraint in the analyzed responses. Negative values of the constraint effect in all the plots indicate the improvement of the resolution results with the inclusion of the horizontal unimodality constraint compared with the classical implementation. Such a positive influence is specially noticeable in the shape modelling of the minor compound, where the constraint effect is found to be statistically significant. This confirms what was already commented in the theory section about the effect of both implementations of the unimodal constraint in minor peaks. The noisy signals associated with these minor constituents are often reduced to narrow peaks when the classical unimodal constraint is applied, whereas the horizontal unimodality preserves much better the shape of the original signal. The constraint effect is not so essential in the shape recovery of major compounds, since their signals detach clearly from the noise and can be more easily modelled. A remarkable positive influence of the constraint is also detected in the normal probability plot associated with the lack of fit. The value of the constraint effect for this response is not statistically significant, but it is clearly negative (i.e. the use of the horizontal unimodality produces a decrease in the lack of fit). This last conclusion is specially important to confirm the goodness of the analyzed constraint in real data sets, where the lack of fit can be determined in

Table 5
Comparison of the lack of fit for several real data sets after ALS application using different constraints

Data sets		Lack of fit (%)	
Resolution	% Minor comp.	1[a]	2[b]
0.1	5	0.93	2.39
0.1	10	0.58	0.86
0.2	1	0.51	1.72
0.3	1	0.21	0.55
0.4	1	0.79	0.80
0.5	0.5	0.23	0.37
0.8	0.5	0.20	0.33
1.0	0.5	0.30	1.91

[a] Constraints applied in the ALS method: non-negativity in concentration profiles and spectra + horizontal unimodality.
[b] Constraints applied in the ALS method: non-negativity in concentration profiles and spectra + classical unimodality.

contrast to the dissimilarities between actual and recovered concentration profiles.

Table 5 includes the lack of fit related to the resolution of several real data sets, already described in a previous section. The lower values obtained for this parameter when the unimodal condition is applied according to the horizontal modality confirm the usefulness of this new implementation. Some apparent inconsistencies can be observed when the values in the table are examined from top to bottom (i.e. comparing systems with the same amount of minor compound, lower lack of fit can be occasionally observed in systems with less chromatographic resolution). These unexpected reversals often occur when experimental data with low concentration of minor compound are used, since the noise patterns and levels in the different samples are not reproducible. Despite this fact, the validity of the comparison between the ALS results obtained using the classical unimodality and the horizontal unimodality is not questionable because the pairs of values compared (i.e. lack of fit applying ALS using both implementations) are referred to pairs of runs of the ALS method on the same data matrix.

4.1.2.2. Localized unimodality. The ALS method has been run forcing horizontal unimodality for the concentration profiles in the experiments with negative constraint level. Localized unimodality has been applied instead when the constraint level is positive. Small departures of the unimodal constraint have been allowed in both varieties (tolerance parameter, $r=1.1$).

Fig. 5 shows the normal probability plots related to the assessment of the localized unimodality constraint. None of the plots shows a significant effect of this constraint in the resolution results, not even a noticeable positive or negative effect. Hence, in its current implementation and for the conditions spanned by the designed experiments, the localized unimodality does not seem to affect the quality of the resolution results. Additional studies revealed that no improvements were obtained when the constraint was only applied to the minor compound.

4.1.2.3. Symmetry. The experiments with a negative constraint level have been resolved applying the horizontal unimodality with a tolerance parameter, $r=1.1$. When the constraint level is positive, ALS is run by using the horizontal unimodality and the symmetry constraint. Small departures from both constraints are allowed (tolerance parameter, $r=1.1$).

Fig. 6 shows the normal probability plots connected with the symmetry constraint. None of the responses is affected significantly by the introduction of the symmetry constraint; so that, no clear modifications can be noticed in the resolution results. The effect of this constraint changes its sign according to the response observed (positive for sigma and negative for the rest of responses); however, this opposed behaviour is not relevant because of the small magnitude of the constraint effect in all the plots examined.

5. Conclusions

The proposal of new constraints related to the experimental features of the data sets has been used as a strategy to improve the solutions coming from the resolution methods. All the constraints presented are related to the modelling of the concentration profiles. Among the constraints, the horizontal unimodality has a more general application and the localized unimodality and symmetry constraint are more focused on the resolution of hyphenated chromatographic systems.

An exhaustive study on a wide span of simulated data sets showed the goodness of the horizontal unim-

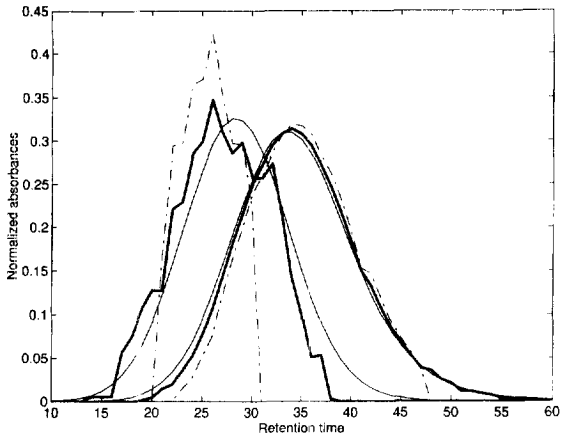

Fig. 7. Comparison of the recovered normalized concentration profiles related to a binary chromatographic system (Rs=0.2, ratio minor : major compound 1 : 100 and *S/N* for the minor compound equal to 20) by using different implementations of the unimodal constraint. True concentration profiles (single lines), profiles recovered using horizontal unimodality (thick lines) and profiles recovered using vertical unimodality (dashed lines). The minor compound is the first eluting.

odality in both the recovery of the shape of the concentration profiles and the error associated with the final solutions. The better quality of the profiles recovered using this kind of unimodality is shown in Fig. 7, where profiles obtained using this new implementation are compared with those obtained using the classical vertical unimodality. The clear decrease of the lack of fit detected in the study performed with simulated data sets allowed the confirmation of the usefulness of this constraint with real data.

Since the horizontal unimodality is exactly equal in concept to the classical unimodality, this better implementation has been applied in the test of the more demanding localized unimodality and symmetry constraints.

Neither the localized unimodality nor the symmetry constraint offered any kind of visible improvement on the resolution results according to the study performed in the present work. Since the real information included in these constraints can be potentially helpful in the modelling of concentration profiles, future research could be oriented to find more effective implementations for these latter constraints. The postulation of new constraints based on different properties of the experimental data could also be explored.

Despite the theoretical validity of the constraints proposed for most of the real chromatographic data sets, the introduction of these constraints in the resolution procedure is always optional and must be supported on the chemical knowledge of the researcher about his data. Evidence of weird behaviours (e.g., fronting phenomena) justify completely the non-application of any of the constraints proposed.

References

[1] R. Tauler, A.K. Smilde and B.R. Kowalski, J. Chemometrics, 9 (1995) 31.
[2] R. Tauler, Chemom. Intell. Lab. Sys., 30 (1995) 133.
[3] R. Tauler and D. Barceló, TrAC, 12 (1993) 319.
[4] W.H. Lawton and E.A. Sylvestre, Technometrics, 13 (1971) 617.
[5] B.G.M. Vandeginste, W. Derks and G. Kateman, Anal. Chim. Acta, 173 (1985) 253.
[6] P.J. Gemperline, Anal. Chem., 58 (1986) 2656.
[7] J. Craig Hamilton and P.J. Gemperline, J. Chemometrics, 4 (1990) 1.
[8] O.M. Kvalheim and Y.Z. Liang, Anal. Chem., 64 (1992) 936.
[9] W. Windig and J. Guilment, Anal. Chem., 63 (1991) 1425.
[10] R. Tauler, A. Izquierdo-Ridorsa, R. Gargallo and E. Casassas, Chemom. Intell. Lab. Sys., 27 (1995) 163.
[11] A. de Juan, G. Fonrodona, R. Gargallo, A. Izquierdo-Ridorsa, R. Tauler and E. Casassas, J. Inorg. Biochem., 63 (1996) 155.
[12] J. Saurina, S. Hernández-Cassou and R. Tauler, Anal. Chem., 67 (1995) 3722.
[13] E. Casassas, R. Tauler and I. Marqués, Macromolecules, 27 (1994) 1729.
[14] S. Lacorte, D. Barceló and R. Tauler, J. Chrom A, 697 (1995) 345.
[15] H. Gampp, M. Maeder, C. Meyer and A.D. Zuberbühler, Talanta, 32 (1985) 1133.
[16] H. Gampp, M. Maeder, C. Meyer and A.D. Zuberbühler, Anal. Chim. Acta, 193 (1987) 287.
[17] A. de Juan, B. van den Bogaert, F. Cuesta Sánchez and D.L. Massart, Chemom. Intell. Lab. Sys., 33 (1996) 133.
[18] F. Cuesta Sánchez, J. Toft, B. van den Bogaert and D.L. Massart, Anal. Chem., 68 (1996) 79.
[19] J.W. Dolan and L.R. Snyder, Troubleshooting LC Systems: A comprehensive approach to troubleshooting LC equipment and separation, Humana Press, US, 1989.
[20] G.E.P. Box, W.G. Hunter and J.S. Hunter, Statistics for Experimenters: An introduction to Design, Data Analysis and Model Building, John Wiley and Sons, US, 1978.
[21] H.R. Keller and D.L. Massart, Anal. Chem., 58 (1993) 471.